The Kantorovich Optimal Transport Problem for Capacities on Finite Spaces

Mario Ghossoub* David Saunders† Kelvin Shuangjian Zhang‡

January 5, 2023

Abstract

In this paper, we study the problem of finding the joint capacity (non-additive measure) with given marginals on a product space that maximizes or minimizes the Choquet integral of a given function. This problem is an analogue for capacities of the classical optimal transport problem for measures. We consider the case in which the marginal spaces both have finitely many points. We provide explicit characterizations of the optimal solutions. We also discuss connections to linear programming, and how the properties of the marginal capacities are reflected by properties of certain capacities in the feasible set. Finally, we discuss potential applications to finance, insurance, and risk management.

1 Introduction

Optimal transport is the subject of a large literature, dating back to the seminal work of Monge (1781) and Kantorovich (1942). Monge (1781) considered the problem of minimizing the total cost (measured using the Euclidean distance between the source and the target) of moving one mass distribution to another among all volume-preserving maps. Kantorovich (1942, 1948) later relaxed this problem by expanding the feasible set to all couplings with given marginal distributions, and developed a duality theory for the relaxed problem. Modern optimal transport is a large and rapidly developing field (e.g. Villani (2008)) with applications to several areas within mathematics (e.g. Rachev and Rüschendorf (1998) and Villani (2003)), and applied fields such as statistics (e.g. Panaretos and Zemel (2022)), economics (e.g. Galichon (2016)), finance (e.g. Henry-Labordère (2017)), and machine learning (e.g. Torres et al. (2021)).

The literature discussed above is based on the assumption that the marginal distributions are additive measures. In many applications, particularly related to the modelling of decision making under ambiguity or vagueness in beliefs, additive measures are replaced by nonadditive monotone set functions, called capacities. See, for example, the foundational work of Schmeidler (1986, 1989), Quiggin (1982, 1993), and Yaari (1987), or the book Grabisch (2016) for several examples. In that context, the standard Lebesgue integral with respect to a measure can be replaced by one of several possible nonlinear generalizations, the most common of which is the Choquet integral with respect to a capacity.[1] In this paper, we consider the problem of finding the joint capacity with given marginals on a product space, which maximizes or minimizes the Choquet integral of a given function (a cost function). This problem is an analogue for capacities of the classical optimal transport problem for measures. We restrict the analysis to the case in which the marginal spaces both have finitely

*Department of Statistics and Actuarial Science, University of Waterloo, mario.ghossoub@uwaterloo.ca
†Department of Statistics and Actuarial Science, University of Waterloo, dsaunders@uwaterloo.ca
‡Department of Statistics and Actuarial Science, University of Waterloo, ks3zhang@uwaterloo.ca

[1] See, for instance, Denneberg (1994) and Marinacci and Montrucchio (2004), for more about capacities and Choquet integration. In addition, note that our use of the word capacity here is distinct from the usage in the literature on optimal transport with capacity constraints, e.g., Korman and McCann (2015); Korman et al. (2015); Pennanen and Perkkiö (2019), where the "capacity constraint" imposes an upper bound on the density of the coupling.

many points. In particular, we show that there exist optimal solutions to both the maximization and minimization problems, that they can be characterized as the (setwise) maximum and minimum capacities in the feasible set, and we provide explicit formulas for these optimal joint capacities in terms of the marginal capacities. We further study properties of the optimal capacities (in particular, non-emptiness of the core) in terms of the corresponding properties of the marginal capacities. We show that, as in the case of measures, the optimal transport problem for capacities can be formulated as a linear program (see Torra (2022) for a related result), and we characterize its dual. Finally, we discuss an application of the optimal transport problem for capacities to risk management; namely, the computation of bounds on risk measures for portfolio losses depending on multiple risk factors with given marginal risk measures.

The literature on the optimal transport problem for capacities described above is thin. Gal and Niculescu (2019) study the problem in the case of capacities on compact metric spaces. Assuming that the cost function c is continuous, they show existence of an optimal joint capacity. Under the further assumption that the marginal capacities are supermodular, and that there exists a supermodular joint capacity, they show the existence of a minimizer when the feasible set is restricted to all supermodular joint capacities, and that optimal supermodular capacities have c-cyclically monotone support. Torra (2022) formulates a version of the optimal transport problem for capacities employing either the Möbius transform (see below), or the “max-plus” transform. He then investigates properties of optimizers and the relationship to standard optimal transport problems for additive measures. Several other authors have studied capacities on product spaces and their properties (e.g. Hendon et al. (1996), Scarsini (1996), Koshevoy (1998), Destercke (2013), and Schmelzer (2015)), with particular attention having been paid to Fubini-type results for Choquet integrals (e.g. Ghirardato (1997), Chateauneuf and Lefort (2008), Bauer (2012), and Wang (2016)).

The remainder of this paper is structured as follows. Section 2 presents definitions and background material needed for the rest of the paper. Section 3 presents a mathematical formulation of the optimal transport problem for capacities, investigates properties of its feasible set, and gives characterizations and explicit formulas for its solution. The existence of the core of the optimal solution and its form are also discussed. Section 4 presents the optimal transport problem for capacities as a linear program, and discusses its dual problem. Section 5 discusses a potential applications in finance and risk management. Finally, Section 6 concludes and discusses possible directions for future research.

2 Definitions and Background Material

In this section, we present some definitions and preliminary remarks, and we set the notation that will be used in the remainder of the paper.

Definition 1. *Let $\mathcal{Z}$ be a nonempty finite set, and let $2^{\mathcal{Z}}$ be the collection of all of its subsets. A function $\gamma : 2^{\mathcal{Z}} \to \mathbb{R}$ is called a capacity if $\gamma(\emptyset) = 0$, and $A \subseteq B$ implies $\gamma(A) \leq \gamma(B)$ for any $A, B \subseteq \mathcal{Z}$. The capacity γ is said to be normalized if $\gamma(\mathcal{Z}) = 1$.*

If γ is a capacity on $\mathcal{Z}$, then its conjugate is defined as $\bar{\gamma}(A) := \gamma(\mathcal{Z}) - \gamma(A^c)$ for all $A \subseteq \mathcal{Z}$. Then $\bar{\gamma}$ is also a capacity, and if γ is normalized then so is $\bar{\gamma}$.

Definition 2. *Let $\mathcal{Z}$ be a nonempty finite set, γ a capacity on $\mathcal{Z}$, and $f : \mathcal{Z} \to \mathbb{R}_+$ be a nonnegative function on $\mathcal{Z}$. The Choquet integral of f with respect to γ is defined as*

$$\gamma(f) := \int_0^\infty \gamma(\{f \geq t\})\, dt.$$

For a function $g : \mathcal{Z} \to \mathbb{R}$ that is not necessarily non-negative, the Choquet integral is defined as

$$\gamma(g) := \int_0^\infty \gamma(\{g \geq t\})\, dt + \int_{-\infty}^0 \left(\gamma(\{g \geq t\}) - \gamma(\mathcal{Z})\right)\, dt.$$

In particular, $\gamma(\mathbf{1}_A) = \gamma(A)$. A significant difference between standard integrals with respect to measures and the Choquet integral is that the Choquet integral is not linear, i.e. we need not have

$\gamma(f+g) = \gamma(f) + \gamma(g)$, or $\gamma(-f) = -\gamma(f)$. As a result, for any fixed capacity γ, the Choquet integral is not homogeneous as a functional. In particular, $\gamma(-f) \neq -\gamma(f)$.[2]

Definition 3. *Let γ be a capacity on $\mathcal{Z}$. The core of γ is the set $\mathcal{C}(\gamma)$ of all measures v on $2^{\mathcal{Z}}$ such that $v(A) \geq \gamma(A)$ for all $A \in 2^{\mathcal{Z}}$ and $v(\mathcal{Z}) = \gamma(\mathcal{Z})$.*

Throughout, we identify measures on any nonempty finite set $\mathcal{Z}$ with vectors $v \in \mathbb{R}^{|\mathcal{Z}|}$ through $v(A) = \sum_{i \in A} v_i$. The core can be empty. For example, consider $\mathcal{Z} = \{z_1, z_2\}$, $\gamma(\{z_1\}) = \gamma(\{z_2\}) = \frac{2}{3}$, and $\gamma(\mathcal{Z}) = 1$. Any $v \in \mathcal{C}(\gamma)$ would have to satisfy $v(\{z_i\}) \geq \frac{2}{3}$, but $v(\mathcal{Z}) = v(\{z_1\}) + v(\{z_2\}) = 1$.

Definition 4. *The Möbius transform of a capacity γ is defined as*

$$m^{\gamma}(A) := \sum_{B \subseteq A} (-1)^{|A \setminus B|} \gamma(B).$$

The Choquet integral of the function f with respect to the capacity γ can be represented in terms of the Möbius transform as follows

$$\begin{aligned} \gamma(f) &= \sum_{A \subseteq \mathcal{X}} m^{\gamma}(A) \bigwedge_{x \in A} f_x = \sum_{A \subseteq \mathcal{X}} \sum_{B \subseteq A} (-1)^{|A \setminus B|} \gamma(B) \bigwedge_{x \in A} f_x \\ &= \sum_{B \subseteq \mathcal{X}} \gamma(B) \left(\sum_{A \supseteq B} (-1)^{|A \setminus B|} \bigwedge_{x \in A} f_x \right) = \sum_{B \subseteq \mathcal{X}} K_f(B) \gamma(B), \end{aligned} \tag{1}$$

with

$$K_f(B) := \sum_{A \supseteq B} (-1)^{|A \setminus B|} \bigwedge_{x \in A} f_x, \tag{2}$$

where $f_x = f(x)$ and $\bigwedge_{x \in A} f_x$ represents the minimum of f on A (Grabisch (2016), Theorem 4.95, page 273). See Grabisch (2016) and Marinacci and Montrucchio (2004) for more information about the Möbius transform.

Definition 5. *Let $\mathcal{Z}$ be a nonempty finite set, and let $\mathcal{G} \subseteq 2^{\mathcal{Z}}$ be a collection of subsets containing $\mathcal{Z}$ and the empty set. Suppose that a function $G : \mathcal{G} \to \mathbb{R}_+$ satisfies $G(\emptyset) = 0$, and $G(A) \leq G(B)$ whenever $A, B \in \mathcal{G}$, $A \subseteq B$. The capacity on $\mathcal{Z}$ defined by*

$$G^*(B) := \inf_{\substack{A \in \mathcal{G} \\ A \supseteq B}} G(A), \text{ for all } B \in 2^{\mathcal{Z}},$$

is called the outer envelope of G. The capacity defined by

$$G_*(B) := \sup_{\substack{A \in \mathcal{G} \\ A \subseteq B}} G(A), \text{ for all } B \in 2^{\mathcal{Z}},$$

is called the inner envelope of G. When it is necessary to make $\mathcal{G}$ explicit in the notation, we will write $G^*(B) = G^*(B; \mathcal{G})$ for the outer envelope, and $G_*(B) = G_*(B; \mathcal{G})$ for the inner envelope.

It is easy to see that $G_* \leq G^*$.[3]

Definition 6. *Given nonempty finite sets $\mathcal{X}, \mathcal{Y}$, we define $\mathcal{P}_{\mathcal{X},\mathcal{Y}}$ to be the collection of all subsets of $\mathcal{X} \times \mathcal{Y}$ of the form $A \times B$ with $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Y}$. We define $\mathcal{P}^*_{\mathcal{X},\mathcal{Y}}$ to be the collection of all subsets of $\mathcal{X} \times \mathcal{Y}$ of the form $A \times B$ with $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Y}$, and either $A = \mathcal{X}$ or $B = \mathcal{Y}$ (or both). That is $\mathcal{P}^*_{\mathcal{X},\mathcal{Y}}$ is the collection of all sets either of the form $\mathcal{X} \times B$ with $B \subseteq \mathcal{Y}$ or $A \times \mathcal{Y}$ with $A \subseteq \mathcal{X}$.*

Sets in product spaces and their projections will feature prominently in the optimal solutions of our optimization problems. The notation in the next definition will be convenient.

[2] However, it is *positive homogeneous*, i.e. $\gamma(\lambda f) = \lambda \gamma(f)$ for $\lambda \geq 0$.

[3] Fix $M \subseteq \mathcal{Z}$, and $A, B \in \mathcal{G}$ with $A \subseteq M \subseteq B$. We have that $G(A) \leq G(B)$. Minimizing over B containing M yields $G(A) \leq G^*(M)$, and then maximizing over A contained in M gives that $G_*(M) \leq G^*(M)$.

Definition 7. *For a set $M \subseteq \mathcal{X} \times \mathcal{Y}$, define:*

$$M_{\mathcal{X}} := \{x \in \mathcal{X} : \exists z = (x, y) \in M\}, \quad M_{\mathcal{Y}} := \{y \in \mathcal{Y} : \exists z = (x, y) \in M\},$$
$$\tilde{M}_{\mathcal{X}} := \{x \in \mathcal{X} : (x, y) \in M, \ \forall y \in \mathcal{Y}\}, \quad \tilde{M}_{\mathcal{Y}} := \{y \in \mathcal{Y} : (x, y) \in M, \ \forall x \in \mathcal{X}\}.$$

It is easy to see that $\tilde{M}_{\mathcal{X}} = ((M^c)_{\mathcal{X}})^c$, and $\tilde{M}_{\mathcal{Y}} = ((M^c)_{\mathcal{Y}})^c$.

Definition 8. *Let $k \geq 2$ be an integer. A capacity γ on $\mathcal{Z}$ is called k-monotone if for any sets $A_1, \ldots, A_k \in \mathcal{Z}$,*

$$\gamma\left(\bigcup_{j=1}^{k} A_j\right) \geq \sum_{\substack{J \subseteq \{1,\ldots,k\} \\ J \neq \emptyset}} (-1)^{|J|+1} \gamma\left(\bigcap_{j \in J} A_j\right).$$

The capacity is called k-alternating if the above inequality is reversed. A 2-monotone capacity is called supermodular, while a 2-alternating capacity is called submodular. If γ is k-monotone for all $k \geq 2$, it is called totally monotone, and if it is k-alternating for all $k \geq 2$, it is called totally alternating.

3 The Optimal Transport Problem for Capacities

In this section, we formulate the optimal transport problem for capacities. Once the problem is formulated, we investigate properties of the feasible set. Understanding the lattice structure of the feasible set leads immediately to explicit formulas for the optimizers.

Definition 9. *Let $\mathcal{X}$ and $\mathcal{Y}$ be non-empty finite sets, and let u and v be probability measures on $\mathcal{X}$ and $\mathcal{Y}$, respectively. Denote by $\Pi_a(u, v)$ the set of measures on $\mathcal{X} \times \mathcal{Y}$ that have the marginals u on $\mathcal{X}$ and v on $\mathcal{Y}$. That is,*

$$\Pi_a(u, v) := \Big\{\pi \mid \pi \text{ is a measure on } \mathcal{X} \times \mathcal{Y} \text{ such that } \pi(A \times \mathcal{Y}) = u(A) \text{ and } \pi(\mathcal{X} \times B) = v(B), \text{ for any } A \subseteq \mathcal{X}, B \subseteq \mathcal{Y}. \Big\}$$

Given a function f, the optimal transport minimization problem is:

$$\inf_{\pi \in \Pi_a(u,v)} \pi(f) = \inf_{\pi \in \Pi_a(u,v)} \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} f(x, y)\, \pi(\{(x, y)\}). \tag{3}$$

Similarly, given a function g, the optimal transport maximization problem is:

$$\sup_{\pi \in \Pi_a(u,v)} \pi(g). \tag{4}$$

Both the maximization and minimization problems are linear in π. Because $\Pi_a(u, v)$ is convex and compact, optimal solutions exist, and the set of optimal solutions contains at least one extreme point of the feasible set. For instance, when $|\mathcal{X}| = |\mathcal{Y}|$ and both u and v are uniform measures, by Birkhoff's Theorem there exists an optimal solution supported on $\bigcup_{i=1}^{|\mathcal{X}|} \{(x_i, y_{\sigma(i)})\}$, for some permutation σ.

Definition 10. *Let $\mathcal{X}$ and $\mathcal{Y}$ be nonempty finite sets and π be a capacity on $\mathcal{X} \times \mathcal{Y}$. The marginal capacities of π on $\mathcal{X}$ and $\mathcal{Y}$, respectively, are defined by*

$$\pi_{\mathcal{X}}(A) := \pi(A \times \mathcal{Y}) \quad \text{and} \quad \pi_{\mathcal{Y}}(B) := \pi(\mathcal{X} \times B),$$

for all $A \subseteq \mathcal{X}$, $B \subseteq \mathcal{Y}$.

The notation defined below for the feasible set of the optimal transport problem for capacities is borrowed from Gal and Niculescu (2019).

Definition 11. *Let μ be a capacity on $\mathcal{X}$ and ν a capacity on $\mathcal{Y}$. The set of all capacities π on $\mathcal{X} \times \mathcal{Y}$ such that $\pi_{\mathcal{X}} = \mu$ and $\pi_{\mathcal{Y}} = \nu$ is denoted by $\Pi_{\mathrm{Ch}}(\mu, \nu)$.*

In particular, for two probability measures u and v, $\Pi_a(u,v) \subseteq \Pi_{\mathrm{Ch}}(u,v)$. The proof of the following result is straightforward.

Lemma 3.1. *Let μ and ν be normalized capacities on $\mathcal{X}$ and $\mathcal{Y}$, respectively. Then $\pi \in \Pi_{\mathrm{Ch}}(\mu,\nu)$ if and only if $\bar{\pi} \in \Pi_{\mathrm{Ch}}(\bar{\mu},\bar{\nu})$.*

Given a function $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, consider the analogue of the optimal transport problem on capacity couplings, i.e. finding capacities to maximize or minimize the Choquet integral of f among all capacities in $\Pi_{\mathrm{Ch}}(\mu,\nu)$:

$$\mathcal{L}(f;\Pi_{\mathrm{Ch}}(\mu,\nu)) := \inf_{\pi\in\Pi_{\mathrm{Ch}}(\mu,\nu)} \pi(f) \leq \sup_{\pi\in\Pi_{\mathrm{Ch}}(\mu,\nu)} \pi(f) =: \mathcal{U}(f;\Pi_{\mathrm{Ch}}(\mu,\nu)).$$

We note that, since $\pi(-f) \neq -\pi(f)$ in general, it is worthwhile to develop the theories for the minimum and maximum problems in parallel.

3.1 The Feasible Set and Its Properties

The first thing to observe about the feasible set is that it is nonempty.

Proposition 3.2. *Let μ and ν be normalized capacities on $\mathcal{X}$ and $\mathcal{Y}$ respectively. Then $\Pi_{\mathrm{Ch}}(\mu,\nu) \neq \emptyset$.*

Proof. Define the function $G : \mathcal{P}_{\mathcal{X},\mathcal{Y}} \to \mathbb{R}_+$ by $G(A \times B) := \mu(A)\cdot\nu(B)$ for $A \times B \in \mathcal{P}_{\mathcal{X},\mathcal{Y}}$ with $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Y}$. It is easy to verify that both G_* and G^* are in $\Pi(\mu,\nu)$. □

We note that we could have used $\mathcal{P}^*_{\mathcal{X},\mathcal{Y}}$ in place of $\mathcal{P}_{\mathcal{X},\mathcal{Y}}$ in the above argument, and reached the same conclusion. Since $\Pi_{\mathrm{Ch}}(\mu,\nu)$ is defined by a finite system of linear equalities and inequalities, and $0 \leq \pi(B) \leq 1$ for any set B, we in fact have the following result.

Proposition 3.3. *Let μ and ν be normalized capacities on $\mathcal{X}$ and $\mathcal{Y}$. Then $\Pi_{\mathrm{Ch}}(\mu,\nu)$ is a compact, convex polyhedron in $\mathbb{R}^{2^{|\mathcal{X}|\cdot|\mathcal{Y}|}}$.*

Remark 1.

- *A capacity γ is called a unanimity game associated with set F if $\gamma(G) = 1$ if $G \supseteq F$, and $\gamma(G) = 0$ otherwise. If μ is the unanimity game associated with $A \subseteq \mathcal{X}$, and ν is the unanimity game associated with $B \subseteq \mathcal{Y}$, then the unanimity game π associated with $A \times B \subseteq \mathcal{X} \times \mathcal{Y}$ is in $\Pi_{\mathrm{Ch}}(\mu,\nu)$.*

- *Suppose that μ is a totally monotone capacity on $\mathcal{X}$ with Möbius transform m^μ, and ν is a totally monotone capacity on $\mathcal{Y}$ with Möbius transform m^ν, then π defined to be the capacity on $\mathcal{X} \times \mathcal{Y}$ with Möbius transform given by*

$$m^\pi(F) = \begin{cases} m^\mu(A)\cdot m^\nu(B), & F = A \times B, A \subseteq \mathcal{X}, B \subseteq \mathcal{Y}; \\ 0, & \text{otherwise.} \end{cases}$$

  is a totally monotone capacity in $\Pi_{\mathrm{Ch}}(\mu,\nu)$.[4] *For further information on this construction, see Walley and Fine (1982), Hendon et al. (1991), Ghirardato (1997), Koshevoy (1998), Bauer (2012), and Destercke (2013). Combining the above argument with Lemma 3.1, it is easy to see that if μ and ν are totally alternating, then there exists a totally alternating capacity $\pi \in \Pi_{\mathrm{Ch}}(\mu,\nu)$.*

- *A possibility measure is a normalized capacity such that $\gamma(A \cup B) = \max(\gamma(A), \gamma(B))$, from which one can show that $\gamma(A) = \max_{z\in A}\gamma(\{z\})$ (and by normalization, there must exist z such that $\gamma(\{z\}) = 1$). If μ and ν are possibility measures, then $\pi(A) := \max_{(x,y)\in A}\mu(\{x\})\cdot\nu(\{y\})$ defines a possibility measure in $\Pi_{\mathrm{Ch}}(\mu,\nu)$. The conjugate of a possibility measure is called a necessity measure (which satisfies $\gamma(A \cap B) = \min(\gamma(A), \gamma(B))$). Again, using Lemma 3.1 one can show that if μ and ν are necessity measures, then there exists a necessity measure $\pi \in \Pi_{\mathrm{Ch}}(\mu,\nu)$.*

[4] It should be noted that if μ and ν are capacities, this construction does not in general result in a capacity. A counterexample is given by $\mathcal{X} = \mathcal{Y} = \{0,1\}$, $\mu = \nu$, with $\mu(\emptyset) = 0, \mu(\{0\}) = \mu(\{1\}) = 0.7, \mu(\mathcal{X}) = 1$ (Dyckerhoff (2022)).

A capacity is said to be balanced if its core is nonempty. The next result demonstrates that there exists a balanced $\pi \in \Pi_{\mathrm{Ch}}(\mu, \nu)$ if and only if both μ and ν are balanced.

Proposition 3.4. *Let μ and ν be normalized capacities on nonempty finite sets $\mathcal{X}$ and $\mathcal{Y}$, respectively. Then the following are equivalent.*

- *Both μ and ν have nonempty cores (i.e. $\mathcal{C}(\mu) \neq \emptyset$ and $\mathcal{C}(\nu) \neq \emptyset$).*
- *There exists $\pi \in \Pi_{\mathrm{Ch}}(\mu, \nu)$ with a nonempty core.*

Proof. Suppose that $u \in \mathcal{C}(\mu)$ and $v \in \mathcal{C}(\nu)$. Define a measure w on $\mathcal{X} \times \mathcal{Y}$ by $w(\{(x, y)\}) := u(\{x\})v(\{y\})$ and additivity. Further, define $G : \mathcal{P}_{\mathcal{X},\mathcal{Y}} \to \mathbb{R}_+$ by $G(A \times B) := \mu(A) \cdot \nu(B)$ for $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Y}$, and take $\pi = G_* \in \Pi_{\mathrm{Ch}}(\mu, \nu)$. It is easy to see that $\pi(\mathcal{X} \times \mathcal{Y}) = w(\mathcal{X} \times \mathcal{Y})$. Let $M \subseteq \mathcal{X} \times \mathcal{Y}$, and consider $K = A \times B \in \mathcal{P}_{\mathcal{X},\mathcal{Y}}$, $K \subseteq M$. Then:

$$G(K) = \mu(A)\nu(B) \leq \sum_{x\in A}\sum_{y\in B} u(\{x\})v(\{y\}) = \sum_{z=(x,y)\in K} w(\{(x, y)\}) \leq \sum_{z=(x,y)\in M} w(\{(x, y)\}) = w(M).$$

This implies that $\pi(M) = G_*(M) \leq w(M)$, for all $M \subseteq \mathcal{X} \times \mathcal{Y}$. Therefore, $w \in \mathcal{C}(\pi)$.

Conversely, let $\pi \in \Pi(\mu, \nu)$ and $w \in \mathcal{C}(\pi)$, and define for $y \in \mathcal{Y}$, $v(\{y\}) := \sum_{x\in\mathcal{X}} w(\{x, y\})$. With $B \subseteq \mathcal{Y}$, we have

$$v(B) = \sum_{y\in B} v(\{y\}) = \sum_{x\in\mathcal{X},y\in B} w(\{x, y\}) = w(\mathcal{X} \times B) \geq \pi(\mathcal{X} \times B) = \nu(B),$$

with equality when $B = \mathcal{Y}$, and therefore $v \in \mathcal{C}(\nu) \neq \emptyset$. The same argument yields $\mathcal{C}(\mu) \neq \emptyset$. □

Remark 2. *It should be noted that there can exist capacities μ on $\mathcal{X}$ and ν on $\mathcal{Y}$ with nonempty cores and an element $\pi \in \Pi(\mu, \nu)$ with an empty core. Consider $\mathcal{X} = \{x_1, x_2\}$, $\mathcal{Y} = \{y_1, y_2\}$, and take μ and ν to be probability measures on $\mathcal{X}$ and $\mathcal{Y}$ respectively, giving equal weight to each point. Define $\pi \in \Pi(\mu, \nu)$ to give value zero to the empty set, 1 to $\mathcal{X} \times \mathcal{Y}$, $\frac{1}{4}$ to any subset consisting of a single point, $\frac{1}{2}$ to any subset consisting of two points, and $\frac{7}{8}$ to any subset consisting of three points. Any element $w \in \mathcal{C}(\pi)$ would have to satisfy $w(\{(x_1, y_1)\}) \geq \frac{1}{4}$, and $w(\mathcal{X} \times \mathcal{Y} \setminus \{(x_1, y_1)\}) \geq \frac{7}{8}$, and thus $w(\mathcal{X} \times \mathcal{Y}) \geq \frac{9}{8} > 1$, contradicting $w(\mathcal{X} \times \mathcal{Y}) = \pi(\mathcal{X} \times \mathcal{Y}) = 1$.*

3.2 Lattice Structure of the Feasible Set and Characterization of the Optimal Solutions

If we think of normalized capacities on $\mathcal{Z}$ as functions on the collection of subsets $2^{\mathcal{Z}}$, then given two capacities γ and π, we can define, for $A \subseteq \mathcal{Z}$:

$$(\pi \wedge \gamma)(A) := \min(\pi(A), \gamma(A)), \quad (\pi \vee \gamma)(A) := \max(\pi(A), \gamma(A)).$$

With these definitions, $\pi\wedge\gamma$ and $\pi\vee\gamma$ are both capacities, and the collection of all normalized capacities is a bounded distributive lattice, with largest element giving value 1 to all nonempty sets, and smallest element giving value 0 to all sets except $\mathcal{Z}$, which has value 1.[5]

Since all capacities in $\Pi_{\mathrm{Ch}}(\mu, \nu)$ have the same values for sets of the form $A \times \mathcal{Y}$, for $A \subseteq \mathcal{X}$, and $\mathcal{X} \times B$, for $B \subseteq \mathcal{Y}$, we have that $\Pi_{\mathrm{Ch}}(\mu, \nu)$ is a distributive sublattice. Furthermore, $\Pi_{\mathrm{Ch}}(\mu, \nu)$ is bounded (as a lattice) with maximum and minimum elements given by taking setwise maxima and minima:

$$\pi^*(A) = \sup_{\pi\in\Pi_{\mathrm{Ch}}(\mu,\nu)} \pi(A) \quad \text{and} \quad \pi_*(A) = \inf_{\pi\in\Pi_{\mathrm{Ch}}(\mu,\nu)} \pi(A).$$

The next result follows almost immediately from the definition of the Choquet integral.

[5] We note that there is another way of defining lattice operations on capacities, involving setwise maxima and minima of their Möbius transforms. See Grabisch (2016) or Marinacci and Montrucchio (2004) for details.

Theorem 3.5. *For $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, and π_* and π^* described above, we have*

$$\min_{\pi \in \Pi_{\mathrm{Ch}}(\mu,\nu)} \pi(f) = \pi_*(f) \quad \text{and} \quad \max_{\pi \in \Pi_{\mathrm{Ch}}(\mu,\nu)} \pi(f) = \pi^*(f).$$

Proof. We first verify that both π_* and π^* are indeed feasible. Note that if $N = A \times \mathcal{Y}$ for $A \subseteq \mathcal{X}$, then $\pi(N) = \mu(A)$ for all $\pi \in \Pi_{\mathrm{Ch}}(\mu,\nu)$, and therefore $\pi^*(N) = \pi_*(N) = \mu(A)$. Similarly, if $N = \mathcal{X} \times B$ with $B \subseteq \mathcal{Y}$, then $\pi^*(N) = \pi_*(N) = \nu(B)$. Furthermore, by their definitions, both π^* and π_* are non-negative non-decreasing set functions, i.e. capacities. In other words, we have that $\pi^*, \pi_* \in \Pi_{\mathrm{Ch}}(\mu,\nu)$.

Now, by the definition in (3.2), π_* and π^* the achieve set-wise infimum and supremum among $\Pi_{\mathrm{Ch}}(\mu,\nu)$, respectively. Let $\pi \in \Pi_{\mathrm{Ch}}(\mu,\nu)$. Then:

$$\begin{aligned}
\pi(f) &= \int_0^\infty \pi(\{f \geq t\})\, dt + \int_{-\infty}^0 (\pi(\{f \geq t\}) - \pi(\mathcal{Z}))\, dt \\
&= \int_0^\infty \pi(\{f \geq t\})\, dt + \int_{-\infty}^0 (\pi(\{f \geq t\}) - 1)\, dt \\
&\geq \int_0^\infty \pi_*(\{f \geq t\})\, dt + \int_{-\infty}^0 (\pi_*(\{f \geq t\}) - 1)\, dt = \pi_*(f).
\end{aligned}$$

The proof for π^* is similar. ∎

It is possible to find explicit expressions for π_* and π^*.

Theorem 3.6. *For any $N \subseteq \mathcal{X} \times \mathcal{Y}$,*

$$\begin{aligned}
\pi_*(N) &= \max\left(\mu(\tilde{N}_{\mathcal{X}}), \nu(\tilde{N}_{\mathcal{Y}})\right), \\
\pi^*(N) &= \min\left(\mu(N_{\mathcal{X}}), \nu(N_{\mathcal{Y}})\right).
\end{aligned}$$

Proof. Define $G : \mathcal{P}^*_{\mathcal{X},\mathcal{Y}} \to \mathbb{R}$ by the following

$$G(M) := \begin{cases} \mu(A), & \text{if } M = A \times \mathcal{Y}; \\ \nu(B), & \text{if } M = \mathcal{X} \times B. \end{cases}$$

Let $G*$ and G_* be the outer and inner envelope of G as defined in Definition 5 with $\mathcal{G} = \mathcal{P}^*_{\mathcal{X},\mathcal{Y}}$. From the monotonicity of μ on $2^{\mathcal{X}}$ (with the inclusion order), it is not hard to see that, for any $N \in \mathcal{P}^*_{\mathcal{X},\mathcal{Y}}$ with $N = A \times \mathcal{Y}$, one has $G^*(N) = G_*(N) = \mu(A)$. Similarly, for any $N = \mathcal{X} \times B$ with $B \subseteq \mathcal{Y}$, we have $G^*(N) = G_*(N) = \nu(B)$. By definition, G^* and G_* are clearly non-negative and non-decreasing, so $G^*, G_* \in \Pi_{\mathrm{Ch}}(\mu,\nu)$.

For any $N \subseteq \mathcal{X} \times \mathcal{Y}$, $\tilde{N}_{\mathcal{X}} \times \mathcal{Y} \subseteq N \subseteq N_{\mathcal{X}} \times \mathcal{Y}$ and $\mathcal{X} \times \tilde{N}_{\mathcal{Y}} \subseteq N \subseteq \mathcal{X} \times N_{\mathcal{Y}}$. Therefore, $G^*(N) \leq \min(\mu(N_{\mathcal{X}}), \nu(N_{\mathcal{Y}}))$ and $G_*(N) \geq \max(\mu(\tilde{N}_{\mathcal{X}}), \nu(\tilde{N}_{\mathcal{Y}}))$. If $N \subseteq A \times \mathcal{Y}$, then $N_{\mathcal{X}} \subseteq A$, and if $A' \times \mathcal{Y} \subseteq N$, then $A' \subseteq \tilde{N}_{\mathcal{X}}$. The monotonicity of μ and ν then imply that

$$\begin{aligned}
G^*(N) &= \min(\mu(N_{\mathcal{X}}), \nu(N_{\mathcal{Y}})), \\
G_*(N) &= \max(\mu(\tilde{N}_{\mathcal{X}}), \nu(\tilde{N}_{\mathcal{Y}})).
\end{aligned}$$

To complete the proof, we will show that $\pi_* = G_*$ and $\pi^* = G^*$. For any $\pi \in \Pi_{\mathrm{Ch}}(\mu,\nu)$ and $N \subseteq \mathcal{X} \times \mathcal{Y}$, the relation $\tilde{N}_{\mathcal{X}} \times \mathcal{Y} \subseteq N \subseteq N_{\mathcal{X}} \times \mathcal{Y}$ implies that

$$\mu(\tilde{N}_{\mathcal{X}}) = \pi(\tilde{N}_{\mathcal{X}} \times \mathcal{Y}) \leq \pi(N) \leq \pi(N_{\mathcal{X}} \times \mathcal{Y}) = \mu(N_{\mathcal{X}}),$$

and $\mathcal{X} \times \tilde{N}_{\mathcal{Y}} \subseteq N \subseteq \mathcal{X} \times N_{\mathcal{Y}}$ implies that

$$\nu(\tilde{N}_{\mathcal{Y}}) = \pi(\mathcal{X} \times \tilde{N}_{\mathcal{Y}}) \leq \pi(N) \leq \pi(\mathcal{X} \times N_{\mathcal{Y}}) = \nu(N_{\mathcal{Y}}).$$

Therefore,

$$G_*(N) = \max(\mu(\tilde{N}_{\mathcal{X}}), \nu(\tilde{N}_{\mathcal{Y}})) \leq \pi(N) \leq \min(\mu(N_{\mathcal{X}}), \nu(N_{\mathcal{Y}})) = G^*(N).$$

This implies, $G_* \leq \pi_*$ and $\pi^* \leq G^*$. The equalities hold because $G_*, G^* \in \Pi_{\mathrm{Ch}}(\mu,\nu)$. ∎

Remark 3. *If we explicitly include the dependence of the optimizers on the marginal capacities, i.e. when given* μ, ν *write* $\pi_*(\cdot;\mu,\nu)$ *and* $\pi^*(\cdot;\mu,\nu)$ *for the smallest and largest elements of* $\Pi_{\mathrm{Ch}}(\mu,\nu)$*, then it is easy to show that* $\bar{\pi}_*(\cdot;\mu,\nu) = \pi^*(\cdot;\bar{\mu},\bar{\nu})$ *and* $\bar{\pi}^*(\cdot;\mu,\nu) = \pi_*(\cdot;\bar{\mu},\bar{\nu})$*.*

Remark 4.

- *Suppose that* μ *is the unanimity game associated with* $A \subseteq \mathcal{X}$ *and* ν *is the unanimity game associated with* $B \subseteq \mathcal{Y}$*, and* $N \subseteq \mathcal{X} \times \mathcal{Y}$*. Then* $\pi_*(N) = 1$ *if either* $A \times \mathcal{Y} \subseteq N$ *or* $\mathcal{X} \times B \subseteq N$*, and zero otherwise. On the other hand,* $\pi^*(N) = 1$ *if for all* $x_0 \in A$ *there exists* $y(x_0) \in \mathcal{Y}$ *such that* $(x_0, y(x_0)) \in N$ and *for all* $y_0 \in B$ *there exists* $x(y_0) \in \mathcal{X}$ *such that* $(x(y_0), y_0) \in N$*, and* $\pi^*(N) = 0$ *otherwise.*
- *Suppose that* μ *and* ν *are possibility measures, and define* $M : \mathcal{X} \times \mathcal{Y} \to [0,1]$ *by* $M(x,y) := \max(\mu(\{x\}), \nu(\{y\}))$*. Then given* $N \subseteq \mathcal{X} \times \mathcal{Y}$*,*

  $$\pi_*(N) = \max(\max_{x \in \tilde{N}_{\mathcal{X}}} \mu(\{x\}), \max_{y \in \tilde{N}_{\mathcal{Y}}} \nu(\{y\})) = \max_{(x,y) \in \tilde{N}_{\mathcal{X}} \times \tilde{N}_{\mathcal{Y}}} M(x,y).$$

  Define $m : \mathcal{X} \times \mathcal{Y} \to [0,1]$ *by* $m(x,y) := \min(\mu(\{x\}), \nu(\{y\}))$*, then*

  $$\pi^*(N) = \min(\max_{x \in N_{\mathcal{X}}} \mu(\{x\}), \max_{y \in N_{\mathcal{Y}}} \nu(\{y\})) = \max_{(x,y) \in N_{\mathcal{X}} \times N_{\mathcal{Y}}} m(x,y).$$

  When μ *and* ν *are necessity measures, then* π_* *and* π^* *can be calculated using the previous remark.*

Consider $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. For a fixed $x \in \mathcal{X}$, define

$$f_y(x) := \min\{f(x,y) : y \in \mathcal{Y}\} \quad \text{and} \quad f^y(x) := \max\{f(x,y) : y \in \mathcal{Y}\},$$

with $f_x, f^x : \mathcal{Y} \to \mathbb{R}$ defined similarly. Then

$$\begin{aligned} \widetilde{\{f \geq t\}}_{\mathcal{X}} &= \{x \in \mathcal{X} : (x,y) \in \{f \geq t\} \ \forall y \in \mathcal{Y}\} \\ &= \{x \in \mathcal{X} : \min_{y \in \mathcal{Y}} f(x,y) \geq t\} = \{f_y \geq t\}. \end{aligned}$$

Similarly $\widetilde{\{f \geq t\}}_{\mathcal{Y}} = \{f_x \geq t\}$, and therefore

$$\pi_*(\{f \geq t\}) = \max(\mu(\{f_y \geq t\}), \nu(\{f_x \geq t\})),$$

and

$$\pi_*(f) = \int_0^\infty \max(\mu(\{f_y \geq t\}), \nu(\{f_x \geq t\}))\,dt + \int_{-\infty}^0 (\max(\mu(\{f_y \geq t\}), \nu(\{f_x \geq t\})) - 1)\,dt,$$

using the fact that we have assumed μ and ν to be normalized.

Using a similar argument,

$$\begin{aligned} \{f \geq t\}_{\mathcal{X}} &= \{x \in \mathcal{X} : \exists y \in \mathcal{Y}, f(x,y) \geq t\} \\ &= \{x \in \mathcal{X} : \max_{y \in \mathcal{Y}} f(x,y) \geq t\} = \{f^y \geq t\}, \end{aligned}$$

and $\{f \geq t\}_{\mathcal{Y}} = \{f^x \geq t\}$. Thus,

$$\pi^*(\{f \geq t\}) = \min(\mu(\{f^y \geq t\}), \nu(\{f^x \geq t\})),$$

and

$$\pi^*(f) = \int_0^\infty \min(\mu(\{f^y \geq t\}), \nu(\{f^x \geq t\}))\,dt + \int_{-\infty}^0 (\min(\mu(\{f^y \geq t\}), \nu(\{f^x \geq t\})) - 1)\,dt.$$

To conclude, we have

$$
\begin{aligned}
\mathcal{L}(f;\Pi_{\mathrm{Ch}}(\mu,\nu)) &= \min_{\pi\in\Pi_{\mathrm{Ch}}(\mu,\nu)} \pi(f) = \pi_*(f) \\
&= \int_0^\infty \max(\mu(\{f_y \geq t\}),\nu(\{f_x \geq t\}))\,dt + \int_{-\infty}^0 (\max(\mu(\{f_y \geq t\}),\nu(\{f_x \geq t\})) - 1)\,dt \\
&\leq \int_0^\infty \min(\mu(\{f^y \geq t\}),\nu(\{f^x \geq t\}))\,dt + \int_{-\infty}^0 (\min(\mu(\{f^y \geq t\}),\nu(\{f^x \geq t\})) - 1)\,dt \\
&= \pi^*(f) = \max_{\pi\in\Pi_{\mathrm{Ch}}(\mu,\nu)} \pi(f) = \mathcal{U}(f;\Pi(\mu,\nu)).
\end{aligned}
$$

3.3 Balancedness and Cores of the Optimal Solutions

Since $\pi_*(N) \leq \pi(N) \leq \pi^*(N)$, for all $N \subseteq \mathcal{X}\times\mathcal{Y}$ and $\pi \in \Pi_{\mathrm{Ch}}(\mu,\nu)$, we immediately obtain the following result.

Proposition 3.7. *Let μ and ν be normalized capacities on $\mathcal{X}$ and $\mathcal{Y}$, respectively. The following statements regarding the cores hold.*

(a) If $\mathcal{C}(\pi^*) \neq \emptyset$, then $\mathcal{C}(\pi) \neq \emptyset$ for all $\pi \in \Pi_{\mathrm{Ch}}(\mu,\nu)$.

(b) If $\mathcal{C}(\pi_*) = \emptyset$, then $\mathcal{C}(\pi) = \emptyset$ for all $\pi \in \Pi_{\mathrm{Ch}}(\mu,\nu)$.

(c) In particular, $\mathcal{C}(\pi_*) \neq \emptyset$ iff $\mathcal{C}(\mu) \neq \emptyset$ and $\mathcal{C}(\nu) \neq \emptyset$.

Proof. Suppose $p \in \mathcal{C}(\pi^*)$, then for any fixed $\pi \in \Pi(\mu,\nu)$ and any $N \subseteq \mathcal{X}\times\mathcal{Y}$, one has $p(N) \geq \pi^*(N) \geq \pi(N)$, with both equalities hold at $N = \mathcal{X}\times\mathcal{Y}$. Therefore, $p \in \mathcal{C}(\pi)$. Using the same argument, one can show part (b). Proposition 3.4 together with part (b) implies (c). □

However, $\mathcal{C}(\pi^*)$ is typically empty, as per the following result.

Proposition 3.8. *Suppose that μ and ν are normalized capacities on $\mathcal{X}$ and $\mathcal{Y}$, respectively, and $|\mathcal{X}| \geq 2$, $|\mathcal{Y}| \geq 2$. Then $\mathcal{C}(\pi^*) = \emptyset$.*

Proof. Let $\{A_1, A_2\}$ and $\{B_1, B_2\}$ be partitions of $\mathcal{X}$ and $\mathcal{Y}$ respectively, and define:

$$N^1 = (A_1 \times B_1) \cup (A_2 \times B_2) \quad\text{and}\quad N^2 = (A_1 \times B_2) \cup (A_2 \times B_1).$$

Then $N^1_{\mathcal{X}} = N^2_{\mathcal{X}} = \mathcal{X}$, $N^1_{\mathcal{Y}} = N^2_{\mathcal{Y}} = \mathcal{Y}$, so that for the disjoint sets N^1 and N^2, $\pi^*(N^1) = \pi^*(N^2) = 1$. □

We can in fact explicitly identify $\mathcal{C}(\pi_*)$ in terms of $\mathcal{C}(\mu)$ and $\mathcal{C}(\nu)$.

Proposition 3.9. *Let μ and ν be normalized capacities on $\mathcal{X}$ and $\mathcal{Y}$, respectively. Then*

$$\mathcal{C}(\pi_*) = \bigcup_{u\in\mathcal{C}(\mu),v\in\mathcal{C}(\nu)} \Pi_a(u,v).$$

Proof. Let $w \in \mathcal{C}(\pi_*)$, and for each fixed $x_0 \in \mathcal{X}$, $y_0 \in \mathcal{Y}$ define $u_w(\{x_0\}) := \sum_{y\in\mathcal{Y}} w(\{x_0,y\})$, and $v_w(\{y_0\}) := \sum_{x\in\mathcal{X}} w(\{x,y_0\})$. Clearly $w \in \Pi_a(u_w,v_w)$. Furthermore, for $A \subseteq \mathcal{X}$, we have

$$u_w(A) = w(A\times\mathcal{Y}) \geq \pi_*(A\times\mathcal{Y}) = \mu(A),$$

since $\pi_* \in \Pi_{\mathrm{Ch}}(\mu,\nu)$. Thus, $u_w \in \mathcal{C}(\mu)$, and similarly $v_w \in \mathcal{C}(\nu)$.

Conversely, suppose that $w \in \Pi_a(u,v)$ with $u \in \mathcal{C}(\mu)$ and $v \in \mathcal{C}(\nu)$. Clearly, $w(\mathcal{X}\times\mathcal{Y}) = u(\mathcal{X}) = \mu(\mathcal{X}) = 1$. Let $N \subseteq \mathcal{X}\times\mathcal{Y}$, and note that $\tilde{N}_{\mathcal{X}}\times\mathcal{Y} \subseteq N$ and $\mathcal{X}\times\tilde{N}_{\mathcal{Y}} \subseteq N$. Then

$$\pi_*(N) = \max(\mu(\tilde{N}_{\mathcal{X}}),\nu(\tilde{N}_{\mathcal{Y}})) \leq \max(u(\tilde{N}_{\mathcal{X}}),v(\tilde{N}_{\mathcal{Y}})) = \max(w(\tilde{N}_{\mathcal{X}}\times\mathcal{Y}),w(\mathcal{X}\times\tilde{N}_{\mathcal{Y}})) \leq w(N).$$

That is, $w \in \mathcal{C}(\pi_*)$. □

Remark 5. *By Corollary 2.23 (ii), page 41 of Grabisch (2016), γ is supermodular iff for every $A \subseteq B \subseteq \mathcal{X} \times \mathcal{Y}$ and $z \notin B$, $\Delta_z \gamma(A) \leq \Delta_z \gamma(B)$, where $\Delta_z \gamma(A) := \gamma(A \cup \{z\}) - \gamma(A)$, and $\Delta_z \gamma(B)$ is defined similarly. It is well-known that if γ is supermodular, then $\mathcal{C}(\gamma) \neq \emptyset$ (Grabisch (2016), Theorem 3.15, page 155).*

Let $\mathcal{X} = \{x_1, x_2, x_3\}$, and $\mathcal{Y} = \{y_1, y_2, y_3\}$, and let μ be the additive (and therefore supermodular) capacity with $\mu(\{x_1\}) = \mu(\{x_2\}) = 0.1$, and $\mu(\{x_3\}) = 0.8$, with ν defined on $\mathcal{Y}$ in the same way. Define:

$$A := \{(x_1, y_2), (x_1, y_3)\} \quad \text{and} \quad B := \{(x_1, y_2), (x_1, y_3), (x_2, y_3), (x_3, y_3)\},$$

and $z := (x_1, y_1)$. Note that $\tilde{A}_{\mathcal{X}} = \emptyset$, $\tilde{A}_{\mathcal{Y}} = \emptyset$, so $\pi_*(A) = 0$. Also, $\widetilde{(A \cup z)}_{\mathcal{X}} = \{x_1\}$, $\widetilde{(A \cup z)}_{\mathcal{Y}} = \emptyset$, so $\Delta_z \pi_*(A) = \pi_*(A \cup z) = \mu(\{x_1\}) = 0.1$. Furthermore, $\tilde{B}_{\mathcal{X}} = \emptyset$, $\tilde{B}_{\mathcal{Y}} = \{y_3\}$, $\widetilde{(B \cup z)}_{\mathcal{X}} = \{x_1\}$, and $\widetilde{(B \cup z)}_{\mathcal{Y}} = \{y_3\}$, so $\pi_*(B) = \pi_*(B \cup z) = \nu(\{y_3\}) = 0.8$, and $\Delta_z \pi_*(B) = 0$. Thus, we conclude that while π_* has a nonempty core, it is not supermodular.

Definition 12. *A capacity γ on $\mathcal{Z}$ is said to be exact if for every $S \in 2^{\mathcal{Z}} \setminus \emptyset$, there exists a core element $p \in \mathcal{C}(\gamma)$ such that $p(S) = \gamma(S)$.*

We have seen that $\mathcal{C}(\pi^*)$ is typically empty, so that π^* will not be exact. In the case when μ and ν are exact, we may ask whether π_* is exact. That is, we define the capacity $\tilde{\pi} \in \Pi_{\mathrm{Ch}}(\mu, \nu)$ by:

$$\tilde{\pi}(N) := \min \left\{ p(N) : p \in \bigcup_{u \in \mathcal{C}(\mu), v \in \mathcal{C}(\nu)} \Pi_a(u, v) \right\}, \text{ for any } N \subseteq \mathcal{X} \times \mathcal{Y},$$

and we ask whether $\pi_* = \tilde{\pi}$.

Remark 6. *In general $\tilde{\pi}$ as defined above need not be either submodular or supermodular. To see this, consider the case $\mathcal{X} = \mathcal{Y} = \{1, 2, \ldots, n\}$ for some $n \geq 3$, with μ and ν being uniform probability measures, and let π' be the conjugate of $\tilde{\pi}$.*[6] *Then*

$$\pi'(A) = 1 - \tilde{\pi}(A^c) = 1 - \min_{p \in \Pi_a(\mu, \nu)} p(A^c) = \max_{p \in \Pi_a(\mu, \nu)} p(A).$$

By Birkhoff's Theorem, the optima $\tilde{\pi}(A)$ (and similarly $\pi'(A)$) is achieved by measures that put mass $\frac{1}{n}$ on points $\{x_i, y_{\sigma(i)}\}$ for some permutation σ. Consider $A_1 = \{(1, 1)\}$, $z = (n, n)$ and $B_1 = \mathcal{X} \times \mathcal{Y} \setminus \{z\}$. Then it is easy to see that $\Delta_z \pi'(A_1) = \frac{2}{n} - \frac{1}{n} = \frac{1}{n}$, while $\Delta_z \pi'(B_1) = 1 - 1 = 0$. Thus $\Delta_z \pi'(A_1) > \Delta_z \pi'(B_1)$, and $A_1 \subseteq B_1$, so π' is not supermodular (and therefore $\tilde{\pi}$ is not submodular, see Grabisch (2016), Theorem 2.20, page 36). On the other hand, consider $A_2 = \{(1, 1)\}$, $B_2 = \{(1, 1), (2, 1)\}$ and $z = (1, 2)$. Then $\Delta_z \pi'(A_2) = 0$, and $\Delta_z \pi'(B_2) = \frac{1}{n}$. We therefore have that $(B_2 \cup \{z\})^c \subseteq (A_2 \cup \{z\})^c$, and $\Delta_z \tilde{\pi}((B_2 \cup \{z\})^c) = \Delta_z \pi'(B_2) > \Delta_z \pi'(A_2) = \Delta_z \tilde{\pi}((A_2 \cup \{z\})^c)$ (Grabisch (2016), page 33, Theorem 2.16). Thus $\tilde{\pi}$ is not supermodular (and π' is not submodular).

Remark 7. *Let $n \geq 2$, $\mathcal{X} = \{1, \ldots, n\}$ and $\mathcal{Y} = \mathcal{X}$, and take μ and ν to be two probability measures on $\mathcal{X}$ that are not equal. Then $\mathcal{C}(\mu) = \{\mu\}$, and $\mathcal{C}(\nu) = \{\nu\}$, so that $\mathcal{C}(\pi_*) = \Pi_a(\mu, \nu)$. Notice that any element of $\Pi_a(\mu, \nu)$ is also in $\Pi_{\mathrm{Ch}}(\mu, \nu)$. $\Pi_a(\mu, \nu)$ is compact, and for any fixed B, $p(B) = \sum_{\{x,y\} \in B} p(\{x, y\})$ is a continuous function on $\Pi_a(\mu, \nu)$ and therefore its minimum is attained. Consider the set $D = \{(1, 1), (2, 2), \ldots, (n, n)\}$ and $M = D^c$. We have that $\tilde{M}_{\mathcal{X}} = \tilde{M}_{\mathcal{Y}} = \emptyset$, and therefore $\pi_*(M) = 0$. Suppose that π_* was exact. There is a $\pi \in \Pi_a(\mu, \nu)$ such that $\pi(M) = 0$. But then π is concentrated on the diagonal D, contradicting the fact that $\mu \neq \nu$. This implies that π_* is not exact.*

[6] We prefer to avoid the cumbersome notation $\tilde{\tilde{\pi}}$.

4 Linear Programming and the Kantorovich Duality for Capacities

In this section, we formulate the optimal transport problem for capacities as a linear program, and we present its dual. Recall that the Choquet integral of f with respect to a capacity γ on $\mathcal{Z}$ can be written as

$$\int f\, d\gamma = \sum_{B \subseteq \mathcal{Z}} K_f(B)\gamma(B),$$

where

$$K_f(B) = \sum_{A \supseteq B} (-1)^{|A \setminus B|} \bigwedge_{x \in A} f_x.$$

While this expression is not linear in f, it is linear in γ, and since the constraints defining $\Pi_{\mathrm{Ch}}(\mu, \nu)$ are all linear (see Proposition 3.3), the problem of minimizing $\pi(f)$ over all $\pi \in \Pi_{\mathrm{Ch}}(\mu, \nu)$ becomes a linear program:

$$\min_{\pi} \sum_{B \subseteq \mathcal{X} \times \mathcal{Y}} K_c(B)\pi(B), \tag{5}$$

$$\begin{aligned} \pi(G \times \mathcal{Y}) &= \mu(G), && \emptyset \neq G \subseteq \mathcal{X}; \\ \pi(\mathcal{X} \times F) &= \nu(F), && \emptyset \neq F \subseteq \mathcal{Y}; \\ \pi(A \cup w) - \pi(A) &\geq 0, && A \subset \mathcal{X} \times \mathcal{Y}, w = \{(x, y)\} \notin A; \\ \pi(\emptyset) &= 0. \end{aligned} \tag{6}$$

(see Grabisch (2016), pages 81-82). Recall that a subset B of $\mathcal{X} \times \mathcal{Y}$ is in $\mathcal{P}^*_{\mathcal{X},\mathcal{Y}}$ if $B = G \times \mathcal{Y}$ for some $G \subseteq \mathcal{X}$ or $B = \mathcal{X} \times F$ for some $F \in \mathcal{Y}$.

The dual of the above linear program is

$$\max_{\hat{\varphi}, \hat{\psi}, \hat{\rho}} \sum_{G \subseteq \mathcal{X}} \hat{\varphi}(G)\mu(G) + \sum_{F \subseteq \mathcal{Y}} \hat{\psi}(F)\nu(F), \tag{7}$$

subject to

$$\begin{aligned} \hat{\varphi}(G) - \sum_{w \notin G \times \mathcal{Y}} \hat{\rho}(G \times \mathcal{Y}, w) + \sum_{w \in G \times \mathcal{Y}} \hat{\rho}((G \times \mathcal{Y}) \setminus \{w\}, w) &= K_c(G \times \mathcal{Y}), \quad \emptyset \neq G \subsetneqq \mathcal{X}; \\ \hat{\psi}(F) - \sum_{w \notin \mathcal{X} \times F} \hat{\rho}(\mathcal{X} \times F, w) + \sum_{w \in \mathcal{X} \times F} \hat{\rho}((\mathcal{X} \times F) \setminus \{w\}, w) &= K_c(\mathcal{X} \times F), \quad \emptyset \neq F \subsetneqq \mathcal{Y}; \\ \hat{\varphi}(\mathcal{X}) + \hat{\psi}(\mathcal{Y}) + \sum_{w} \hat{\rho}((\mathcal{X} \times \mathcal{Y}) \setminus \{w\}, w) &= K_c(\mathcal{X} \times \mathcal{Y}); \\ - \sum_{w \notin B} \hat{\rho}(B, w) + \sum_{w \in B} \hat{\rho}(B \setminus \{w\}, w) &= K_c(B), \quad B \notin \mathcal{P}^*_{\mathcal{X},\mathcal{Y}}; \\ \hat{\rho} &\geq 0. \end{aligned} \tag{8}$$

Let $(\hat{\varphi}_*, \hat{\psi}_*, \hat{\rho}_*)$ be an optimal solution to (7 - 8). Then complementary slackness implies that, for any $(A, w) \in \{(A, w) \in 2^{\mathcal{X} \times \mathcal{Y}} \times (\mathcal{X} \times \mathcal{Y}) : w \notin A\}$,

$$\hat{\rho}_*(A, w)\ (\pi_*(A \cup w) - \pi_*(A)) = 0. \tag{9}$$

Remark 8. *The dual of the maximization problem*

$$\max_{\pi \in \Pi_{\mathrm{Ch}}(\mu, \nu)} \pi(c) \tag{10}$$

is given by

$$\min_{\hat{\varphi},\hat{\psi},\hat{\rho}} \sum_{G\subseteq\mathcal{X}} \hat{\varphi}(G)\mu(G) + \sum_{F\subseteq\mathcal{Y}} \hat{\psi}(F)\nu(F), \tag{11}$$

subject to

$$\begin{aligned}
\hat{\varphi}(G) - \sum_{w\notin G\times\mathcal{Y}} \hat{\rho}(G\times\mathcal{Y}, w) + \sum_{w\in G\times\mathcal{Y}} \hat{\rho}((G\times\mathcal{Y})\setminus\{w\}, w) &= K_c(G\times\mathcal{Y}), \quad \emptyset\neq G\subsetneq\mathcal{X};\\
\hat{\psi}(F) - \sum_{w\notin \mathcal{X}\times F} \hat{\rho}(\mathcal{X}\times F, w) + \sum_{w\in \mathcal{X}\times F} \hat{\rho}((\mathcal{X}\times F)\setminus\{w\}, w) &= K_c(\mathcal{X}\times F), \quad \emptyset\neq F\subsetneq\mathcal{Y};\\
\hat{\varphi}(\mathcal{X}) + \hat{\psi}(\mathcal{Y}) + \sum_{w} \hat{\rho}((\mathcal{X}\times\mathcal{Y})\setminus\{w\}, w) &= K_c(\mathcal{X}\times\mathcal{Y});\\
-\sum_{w\notin B} \hat{\rho}(B, w) + \sum_{w\in B} \hat{\rho}(B\setminus\{w\}, w) &= K_c(B), \quad B\notin\mathcal{P}^*_{\mathcal{X},\mathcal{Y}};\\
\hat{\rho} &\leq 0.
\end{aligned} \tag{12}$$

Suppose that $(\hat{\varphi}^*, \hat{\psi}^*, \hat{\rho}^*)$ *is an optimal solution to (11 - 12). Then by complementary slackness, for any* $(A, w) \in \{(A, w) \in 2^{\mathcal{X}\times\mathcal{Y}} \times (\mathcal{X}\times\mathcal{Y}) : w \notin A\}$,

$$\hat{\rho}^*(A, w)\ (\pi^*(A\cup w) - \pi^*(A)) = 0.$$

Remark 9. *The dual of the minimization Optimal Transport problem is equivalent to the problem*

$$\max_{L_\varphi, L_\psi, \hat{\rho}} \sum_{G\subseteq\mathcal{X}} m^\mu(G) L_\varphi(G) + \sum_{F\subseteq\mathcal{Y}} m^\nu(F) L_\psi(F), \tag{13}$$

$$\begin{aligned}
L_\varphi(A_\mathcal{X}) + L_\psi(A_\mathcal{Y}) + \sum_{D\supseteq A}\sum_{w\in A} \hat{\rho}(D\setminus\{w\}, w) &= \bigwedge_{(x,y)\in A} c(x, y), \quad \emptyset\neq A\subseteq\mathcal{X}\times\mathcal{Y};\\
\hat{\rho} &\geq 0.
\end{aligned} \tag{14}$$

To see this, we will show, by the following change of variables[7]

$$\hat{\varphi}(G) := \sum_{B\supseteq G} (-1)^{|B\setminus G|} L_\varphi(B);$$
$$\hat{\psi}(F) := \sum_{A\supseteq F} (-1)^{|A\setminus F|} L_\psi(A),$$

that the objectives are equal and that the constraints can be derived from each other.

First, the objective function becomes

$$\begin{aligned}
&\sum_{G\subseteq\mathcal{X}} \hat{\varphi}(G)\mu(G) + \sum_{F\subseteq\mathcal{Y}} \hat{\psi}(F)\nu(F)\\
&= \sum_{G\subseteq\mathcal{X}}\sum_{B\supseteq G} (-1)^{|B\setminus G|} L_\varphi(B)\mu(G) + \sum_{F\subseteq\mathcal{Y}}\sum_{A\supseteq F} (-1)^{|A\setminus F|} L_\psi(A)\nu(F)\\
&= \sum_{B\subseteq\mathcal{X}} \left(\sum_{G\subseteq B} (-1)^{|B\setminus G|}\mu(G)\right) L_\varphi(B) + \sum_{A\subseteq\mathcal{Y}} \left(\sum_{F\subseteq A} (-1)^{|A\setminus F|}\nu(F)\right) L_\psi(A)\\
&= \sum_{B\subseteq\mathcal{X}} m^\mu(B) L_\varphi(B) + \sum_{A\subseteq\mathcal{Y}} m^\nu(A) L_\psi(A).
\end{aligned}$$

[7] This corresponds to the situation derived from a set function ξ_φ, where $\hat{\varphi} = m^{\xi_\varphi}$, and $L_\varphi = \check{m}^{\xi_\varphi}$, the co-Möbius transform, with similar conventions for ψ, see (Grabisch, 2016, Table A.2 on page 440).

To see that the constraints are equivalent, notice that the above transformation can be inverted as

$$L_\varphi(G) = \sum_{G' \supseteq G} \hat{\varphi}(G');$$
$$L_\psi(F) = \sum_{F' \supseteq F} \hat{\psi}(F').$$

Furthermore, for any $B \subseteq \mathcal{X} \times \mathcal{Y}$, recall

$$K_c(B) = \sum_{A \supseteq B} (-1)^{|A \setminus B|} \bigwedge_{(x,y) \in A} c(x,y).$$

Using the same inversion formula, we obtain

$$\bigwedge_{(x,y)\in A} c(x,y) = \sum_{B \supseteq A} K_c(B).$$

For any non-empty set $A \subseteq \mathcal{X} \times \mathcal{Y}$, sum all constraints with a right-hand side involving $K_c(B)$ with $B \supseteq A$. The right-hand side term of (8) becomes

$$\sum_{B \supseteq A} K_c(B) = \bigwedge_{(x,y)\in A} c(x,y).$$

The sum of terms on the left-hand side of (8) will yield a sum involving $\hat{\varphi}$, which is

$$\sum_{G' \supseteq A_{\mathcal{X}}} \hat{\varphi}(G') = L_\varphi(A_{\mathcal{X}}),$$

and a sum involving $\hat{\psi}$, which is

$$\sum_{F' \supseteq A_{\mathcal{Y}}} \hat{\psi}(F') = L_\psi(A_{\mathcal{Y}}).$$

Lastly, denoting the sum of all terms involving $\hat{\rho}$ in (8) by S, we obatin

$$S = J_1 + J_2 + J_3 + J_4 + J_5 + J_6 + J_7,$$

where

$$J_1 := -\sum_{\substack{G' \supseteq A_{\mathcal{X}} \\ G' \neq \mathcal{X}}} \sum_{\substack{x \notin G' \\ y \in \mathcal{Y}}} \hat{\rho}(G' \times \mathcal{Y}, (x,y));$$
$$J_2 := \sum_{\substack{G' \supseteq A_{\mathcal{X}} \\ G' \neq \mathcal{X}}} \sum_{\substack{x \in G' \\ y \in \mathcal{Y}}} \hat{\rho}(G' \times \mathcal{Y} \setminus \{(x,y)\}, (x,y));$$
$$J_3 := -\sum_{\substack{F' \supseteq A_{\mathcal{Y}} \\ F' \neq \mathcal{Y}}} \sum_{\substack{y \notin F' \\ x \in \mathcal{X}}} \hat{\rho}(\mathcal{X} \times F', (x,y));$$
$$J_4 := \sum_{\substack{F' \supseteq A_{\mathcal{Y}} \\ F' \neq \mathcal{Y}}} \sum_{\substack{y \in F' \\ x \in \mathcal{X}}} \hat{\rho}(\mathcal{X} \times F' \setminus \{(x,y)\}, (x,y));$$
$$J_5 := \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \hat{\rho}(\mathcal{X} \times \mathcal{Y} \setminus \{(x,y)\}, (x,y));$$
$$J_6 := -\sum_{\substack{B \supseteq A \\ B \notin \mathcal{P}^*_{\mathcal{X},\mathcal{Y}} \\ (x,y) \notin B}} \hat{\rho}(B, (x,y));$$

$$J_7 := \sum_{\substack{B \supseteq A \\ B \notin \mathcal{P}^*_{\mathcal{X},\mathcal{Y}} \\ (x,y)\in B}} \hat{\rho}(B \setminus \{(x,y)\}, (x,y)).$$

By summing the above terms, we obtain

$$\begin{aligned} S &= -\sum_{B\supseteq A}\sum_{w\notin B}\hat{\rho}(B,w) + \sum_{D\supseteq A}\sum_{w\in D}\hat{\rho}(D\setminus\{w\},w) \\ &= -\sum_{B\supseteq A}\sum_{w\notin B}\hat{\rho}(B,w) + \sum_{D\supseteq A}\left[\sum_{w\in A}\hat{\rho}(D\setminus\{w\},w) + \sum_{w\in D\setminus A}\hat{\rho}(D\setminus\{w\},w)\right] \\ &= \sum_{D\supseteq A}\sum_{w\in A}\hat{\rho}(D\setminus\{w\},w). \end{aligned} \tag{15}$$

The last equality comes from the observation that there exists an one-to-one mapping between $\{(B,w) : A \subset B, w \notin B\}$ *and* $\{(D,w) : A \subset D, w \in D\setminus A\}$ *by the map* $D := B \cup \{w\}$*, and thus the first and third terms in the second line of* (15) *cancel out. Therefore, one can derive the equations in* (14) *from those in* (8)*. Similarly, one can also prove the opposite direction by using the above change of variables.*

Remark 10. *By a similar argument, the dual of the maximization Optimal Transport problem is equivalent to*

$$\min_{L_\varphi, L_\psi, \hat{\rho}} \sum_{G\subseteq\mathcal{X}} m^\mu(G)L_\varphi(G) + \sum_{F\subseteq\mathcal{Y}} m^\nu(F)L_\psi(F), \tag{16}$$

$$\begin{aligned} L_\varphi(A_\mathcal{X}) + L_\psi(A_\mathcal{Y}) + \sum_{D\supseteq A}\sum_{w\in A}\hat{\rho}(D\setminus\{w\},w) &= \bigwedge_{(x,y)\in A} c(x,y), \quad \emptyset \neq A \subseteq \mathcal{X}\times\mathcal{Y}; \\ \hat{\rho} &\leq 0. \end{aligned} \tag{17}$$

5 Bounds on Risk Measures for Losses Depending on Multiple Risk Factors with Given Marginal Risk Measures

In this section, we briefly discuss an application of the optimal transport problem for capacities: the computation of bounds on risk measures for portfolios whose losses depend on two sets of risk factors, for which risk measures are known for portfolios that depend on the factors separately, but not jointly.

A risk measure R is a real-valued functional on $\mathcal{L} := L^0(\mathcal{X}\times\mathcal{Y})$, the space of functions on $\mathcal{X}\times\mathcal{Y}$. Common properties that a risk measure may satisfy include:

R.1 (Monotonicity) $R(X) \leq R(Y)$, for all $X, Y \in \mathcal{L}$ such that $X \leq Y$, π-a.s.

R.2 (Positive Homogeneity) $R(\lambda X) = \lambda R(X)$, for all $X \in \mathcal{L}$ and all $\lambda \in \mathbb{R}_+$.

R.3 (Cash Invariance) $R(X + c) = R(X) + c$, for all $X \in \mathcal{L}$ and $c \in \mathbb{R}$.

R.4 (Subadditivity) $R(X + Y) \leq R(X) + R(Y)$ for all $X, Y \in \mathcal{L}$.

R.5 (Convexity) $R(\lambda X + (1-\lambda)Y) \leq \lambda R(X) + (1-\lambda)R(Y)$ for all $X, Y \in \mathcal{L}$ and $\lambda \in [0,1]$.

Subadditivity and positive homogeneity clearly imply convexity. Following Artzner et al. (1999), a risk measure is called coherent if it satisfies the first four of the above axioms. For any fixed capacity γ, the Choquet integral defines a risk measure R_γ by $R_\gamma(f) := \gamma(f)$. Furthermore, it follows immediately from the definition that the risk measure R_γ is monotonic, positively homogeneous, and cash invariant.

Risk measures defined by Choquet integrals have an additional important property. We recall that two functions $f, g : \mathcal{Z} \to \mathbb{R}$ are comonotonic if $(f(z) - f(z')) \cdot (g(z) - g(z')) \geq 0$ for all $z, z' \in \mathcal{Z}$. A risk measure R is said to be comonotonic additive if $R(f+g) = R(f) + R(g)$ whenever f and g are comonotonic. For any capacity γ, it follows from the definition of the Choquet integral that the risk measure R_γ is comonotonic additive. Indeed, a risk measure R is monotone, translation invariant, and comonotonic additive if and only if there is a capacity γ such that $R = R_\gamma$ (Föllmer and Schied (2016), Theorem 4.88, page 258).

Suppose that we are given a risk measure $R_{\mathcal{X}}$ on functions on $\mathcal{X}$, and a risk measure $R_{\mathcal{Y}}$ on functions on $\mathcal{Y}$, both of which are monotone, translation invariant, and comonotonic additive, and therefore represented by Choquet integrals with respect to marginal capacities μ and ν. The spaces $\mathcal{X}$ and $\mathcal{Y}$ could represent, for example, risk factors affecting different lines of business for an insurance company, or sets of risk factors affecting the losses of a portfolio in different ways. For example, in counterparty credit risk, losses depend on the creditworthiness of the counterparties (affected by credit risk factors), and the value of a portfolio of derivative contracts with the counterparties (market risk factors). Models for the distribution and risk of these factors may be available separately, but no jointly calibrated model for all risk factors may be available. Given a portfolio loss function that depends on both the factors $\mathcal{X}$ and $\mathcal{Y}$ (i.e. a function on $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$), the Choquet integrals $\pi_*(f)$ and $\pi^*(f)$ give bounds on the joint portfolio risk. More specifically, as the optimal value of the problem $\max\{\pi(f) : \Pi_{\mathrm{Ch}}(\mu, \nu)\}$, $\pi^*(f)$ gives an upper bound on the risk of f among all monotone, translation invariant, and comonotonic additive risk measures that agree with the given risk measures $R_{\mathcal{X}}$ and $R_{\mathcal{Y}}$ for losses that only depend on the factors $\mathcal{X}$ and $\mathcal{Y}$, respectively.

A related question is the following. Given coherent comonotonic additive risk measures $R_{\mathcal{X}}$ and $R_{\mathcal{Y}}$ defined on functions on $\mathcal{X}$ and $\mathcal{Y}$, respectively, can one find upper and lower bounds on the risk of a loss function defined on $\mathcal{X} \times \mathcal{Y}$? That is, among all coherent and comonotonic additve risk measures $R_{\mathcal{X},\mathcal{Y}}$ defined on functions on $\mathcal{X} \times \mathcal{Y}$, and which agree with $R_{\mathcal{X}}$ and $R_{\mathcal{Y}}$ for functions that depend on only one set of risk factors, what are the largest and smallest values for $R_{\mathcal{X},\mathcal{Y}}(f)$? This question can be considered with a modification of our approach. In particular, it is known that any convex, comonotonic additive, and monotone risk measure R can be represented as $R(f) = \gamma(f)$, for some normalized and submodular capacity γ (and any such γ defines a risk measure with the stated properties). Furthermore, any such convex risk measure is actually positively homogeneous, and therefore coherent (see Föllmer and Schied (2016), Section 4.7 for proofs of these results). Thus, the stated problem is equivalent to being given submodular marginal capacities μ, ν and seeking a *submodular* π in $\Pi_{\mathrm{Ch}}(\mu, \nu)$ that maximizes or minimizes the Choquet integral of a given loss function f. The solutions of our problem presented earlier can be regarded as solutions of the relaxed problem, in which the submodularity constraint is removed. Alternatively, the problem with the submodularity constraint can be formulated as a linear program, as in Section 4. In particular, if we add the (linear) constraints (see Grabisch (2016), Corollary 2.23, page 41)

$$\pi(A \cup \{v, w\}) - \pi(A \cup \{v\}) - \pi(A \cup \{w\}) + \pi(A) \leq 0, \forall A \subseteq \mathcal{X} \times \mathcal{Y}, v, w \notin A$$

to problems (5) and (10), then we maximize and minimize the risk among all coherent, comonotonic additive risk measures. Gal and Niculescu (2019) study this problem (for supermodular capacities, but the problems are equivalent) in the context of capacities on general complete separable metric spaces. Under technical assumptions including the existence of a supermodular capacity in $\Pi_{\mathrm{Ch}}(\mu, \nu)$ given supermodular μ and ν, they show that an optimizer exists, and has c-cyclically monotone support (for a given continuous cost function c). Unfortunately, it remains open in this case whether the feasible set is nonempty, i.e. whether given supermodular capacities μ and ν, there exists a supermodular capacity $\pi \in \Pi_{\mathrm{Ch}}(\mu, \nu)$.

For related work on bounds on risk measures when the marginal distributions of the risk factors are known with certainty, and there is uncertainty on the joint distribution, see Hall et al. (2022), Memartoluie et al. (2017), Ennaji et al. (2022), and the references therein. Related problems of "generalized optimal transport", where the feasible set is $\Pi_a(m, n)$ for two measures m and n, but the objective function is not an expectation, are studied in Pennanen and Perkkiö (2019).

6 Conclusions and Directions for Future Research

In this paper, we studied the optimal transport problem for capacities, that is maximizing or minimizing the Choquet integral of a given function with respect to a capacity on a product space with prescribed marginal capacities. In the case where both marginal spaces consist of finitely many points, we showed that the problem always has an optimal solution, for which we are able to give two characterizations, one by setwise maximization/minimization among all feasible capacities, and the other through an explicit formula involving projections of sets and the marginal capacities. We further investigated the relationship between properties of the marginal capacities and those of the optimizers (or more generally capacities in the feasible set). In particular, we showed that the minimizing capacity π_* is balanced if and only if both marginal capacities are balanced, and we determined its core explicitly in this case. In all but the most trivial cases, the maximizing capacity π^* is not balanced. We further discussed connections with linear programming, showing that the optimal transport problems for capacities are linear programs, and we found their duals explicitly. Applications to problems in finance and insurance were also discussed.

The most obvious direction for generalization is to the case where the marginal spaces $\mathcal{X}$ and $\mathcal{Y}$ are not finite, but instead topological spaces with some structure, say complete separable metric spaces. This is the subject of ongoing work. Characterizing optimal solutions is more challenging in this case, as nontrivial issues regarding measurability of projections arise. Other directions for future work include developing the applications in finance outlined briefly above, and developing efficient numerical methods, particularly in the infinite dimensional case. There are also several questions that could be posed about the structure of the feasible set $\Pi_{\mathrm{Ch}}(\mu, \nu)$ given various assumptions on μ and ν. As an example, several of the results in Gal and Niculescu (2019) depend on the assumption of the existence of a supermodular $\pi \in \Pi_{\mathrm{Ch}}(\mu, \nu)$ given that μ and ν are supermodular. The existing results investigating this condition require strong assumptions for it to hold.[8]

References

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

C. Bauer. Products of non-additive measures: a Fubini-like theorem. *Theory and Decision*, 73:621–647, 2012.

A. Chateauneuf and J.-P. Lefort. Some Fubini theorems on product σ-algebras for non-additive measures. *International Journal of Approximate Reasoning*, 48:686–696, 2008.

Dieter Denneberg. *Non-additive measure and integral*, volume 27. Springer Science & Business Media, 1994.

S. Destercke. Independence and 2-monotonicity: Nice to have, hard to keep. *International Journal of Approximate Reasoning*, 54:478–490, 2013.

R. Dyckerhoff. Personal communication, 2022.

H. Ennaji, Q. Mérigot, L. Nenna, and B. Pass. Robust risk management via multi-marginal optimal transport. *arXiv preprint arXiv:2211.07694*, 2022.

H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*. de Gruyter, Berlin, fourth edition, 2016.

S.G. Gal and C.P. Niculescu. Kantorovich's mass transport problem for capacities. *Proceedings of the Romanian Academy, Series A*, 20(4):337–345, 2019.

[8] If μ and ν are both totally monotone, the construction using products of Möbius transforms in Remark 1 yields a totally monotone product capacity. Ghirardato (1997) shows that if μ is additive, and ν is supermodular, then there exists a unique supermodular capacity $\pi \in \Pi_{\mathrm{Ch}}(\mu, \nu)$ satisfying $\pi(A \times B) = \mu(A)\nu(B) \forall A \in \mathcal{X}, B \in \mathcal{Y}$.

A. Galichon. *Optimal Transport Methods in Economics.* Princeton University Press, Princeton, 2016.

P. Ghirardato. On independence for non-additive measures, with a Fubini theorem. *Journal of Economic Theory*, 73(2):261–291, 1997.

M. Grabisch. *Set Functions, Games, and Capacities in Decision Making.* Springer, Switzerland, 2016.

J. Hall, M. Ghossoub, and D. Saunders. Maximum spectral measures of risk with given risk factor marginal distributions. *Mathematics of Operations Research*, 2022. To appear.

E. Hendon, H.J. Jacobsen, B. Sloth, and T. Tran. The product of capacities and lower probabilities. Technical report, mimeo, University of Copenhagen, 1991.

E. Hendon, H.J. Jacobsen, B. Sloth, and T. Tranæs. The product of capacities and belief functions. *Mathematical Social Sciences*, 32:95–108, 1996.

P. Henry-Labordère. *Model-Free Hedging: A Martingale Optimal Transport Viewpoint.* CRC Press, Boca Raton, 2017.

L.V. Kantorovich. On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, 37: 199–201, 1942.

L.V. Kantorovich. On a problem of Monge (in russian). *Uspekhi Math. Nauk.*, 3:225–226, 1948.

J. Korman and R. McCann. Optimal transportation with capacity constraints. *Transactions of the American Mathematical Society*, 367(3):1501–1521, 2015.

J. Korman, R. McCann, and C. Seis. Dual potentials for capacity constrained optimal transport. *Calculus of Variations and Partial Differential Equations*, 54(1):573–584, 2015.

G.A. Koshevoy. Distributive lattices and products of capacities. *Journal of Mathematical Analysis and Applications*, 219:427–441, 1998.

M. Marinacci and L. Montrucchio. Introduction to the mathematics of ambiguity. *Uncertainty in Economic Theory, I. Gilboa (ed.)*, pages 46–107, 2004.

A. Memartoluie, D. Saunders, and T. Wirjanto. Wrong-way risk bounds in counterparty credit risk management. *Journal of Risk Management in Financial Institutions*, 10(2):150–163, 2017.

G. Monge. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704, 1781.

V.M. Panaretos and Y. Zemel. *An Invitation to Statistics in Wasserstein Space.* Springer, Cham, 2022.

T. Pennanen and A.-P. Perkkiö. Convex duality in nonlinear optimal transport. *Journal of Functional Analysis*, 277(14):1029–1060, 2019.

J. Quiggin. A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4): 323–343, 1982.

J. Quiggin. *Generalized Expected Utility Theory - The Rank-Dependent Model.* Kluwer Academic Publishers, 1993.

S.T. Rachev and L. Rüschendorf. *Mass Transportation Problems.* Springer, New York, 1998.

M. Scarsini. Copulae of capacities on product spaces. In L. Rüschendorf, B. Schweizer, and M.D. Taylor, editors, *Distributions with Fixed Marginals and Related Topics*, volume 28 of *Institute of Mathematical Statistics Lecture Notes*, pages 307–318. Institute of Mathematical Statistics, Hayward, 1996.

D. Schmeidler. Integral representation without additivity. *Proceedings of the American Mathematical Society*, 97(2):255–261, 1986.

D. Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, pages 571–587, 1989.

B. Schmelzer. Joint distributions of random sets and their relation to copulas. *International Journal of Approximate Reasoning*, 65:59–69, 2015.

V. Torra. The transport problem for non-additive measures. *arXiv preprint arXiv:2211.12150v2*, 2022.

L.C. Torres, L.M. Pereira, and M.H. Amini. A survey on optimal transport for machine learning: Theory and applications. *arXiv preprint arXiv:2106.01963*, 2021.

C. Villani. *Topics in Optimal Transportation.* American Mathematical Society, Providence, 2003.

C. Villani. *Optimal Transport, Old and New.* Springer, Berlin, 2008.

P. Walley and T.L. Fine. Towards a frequentist theory of upper and lower probability. *Annals of Statistics*, 10(3):741–761, 1982.

H. Wang. Fubini theorems for capacities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 24(6):901–916, 2016.

M. Yaari. The dual theory of choice under risk. *Econometrica*, pages 95–115, 1987.